

Enhanced Privacy and Secure Data Mining Protocol For Horizontally Distributed Databases with Cost Effectiveness

Ashish S.Malapure

Research Student,
Government College of Engineering,
Aurangabad, India.

Vivek Kshirsagar

Associate professor,
Government College of Engineering,
Aurangabad, India.

Abstract:

Data mining is the most fast growing area today which is used to extract important knowledge from large data collections but often these collections are divided among several parties. Privacy liability may prevent the parties from directly sharing the data and some types of information about the data. In this project we propose a protocol for secure mining of association rules in horizontally distributed databases. The current integral protocol is that of Kantarcioglu and Clifton well known as K&C protocol. This protocol is based on an unsecured distributed version of the Apriori algorithm named as Fast Distributed Mining (FDM) algorithm of Cheung et al. The main ingredients in our protocol are two novel secure multi-party algorithms one that computes the union of private subsets that each of the interacting players hold and another that tests the whether an element held by one player is included in a subset held by another. This protocol offers enhanced privacy with respect to the earlier protocols. In addition, it is not complicated and is importantly more effectual in terms of communication cost, communication rounds and computational cost.

Keywords: Security, Privacy, Data Mining, Frequent Item sets, Association Rules, multi-party

INTRODUCTION:

Data mining can extract important knowledge from large data collections but sometimes these collections are split among various parties. Privacy liability may pre-vent the parties from directly sharing the data, and some types of information about the data. Data mining technology has become prominent as a means of identifying patterns and trends from large quantities of data.

Data mining and data warehousing co-jointly: most popular tools operate by gathering all data into a central site then running an algorithm against that data. However, privacy liability can prevent building a centralized warehouse data may be distributed among several custodians none of which are allowed to transfer their data to another site. In Horizontally partitioned database there are several players that hold homogeneous database. The goal is to find all association rules with support at least s and confidence at least c , for some given minimal support size s and confidence level c , that hold in the unified database, while minimizing the information disclosed about the private databases held by those players. That goal defines a problem of secure multi-party computation.

If there existed a trusted third party, the players could surrender to him their inputs and he would perform the function evaluation and send to them the resulting output. In the absence of such a trusted third party, it is needed to devise a protocol that the players can run on their own in order to arrive at the required output y . Such a protocol is considered perfectly secure if no player can learn from his view of the protocol more than what he would have learnt in the idealized setting where the computation is carried out by a trusted third party. In previous year various techniques are applied for secure mining of association rules in horizontally partitioned database. These approaches use various techniques such as data perturbation, homo-morphic encryption, keyword search and oblivious pseudorandom functions etc. These privacy preserving approaches are inefficient due to

- Homo-morphic encryption
- Higher computational cost

- In some of the techniques data owner tries to hide data from data miner.

Our proposed protocol based on two novel secure multiparty algorithms using these algorithms the protocol provides enhanced privacy, security and efficiency as it uses commutative encryption. In this project we propose a protocol for secure mining of association rules in horizontally distributed database. This protocol is based on: FDM Algorithm which is an unsecured distributed version of the Apriori algorithm. In our protocol two secure multiparty algorithms are involved:

1. Computes the union of private subsets that each interacting players hold.
2. Tests the inclusion of an element held by one player in subset held by another.

In Horizontally partitioned database there are several players that hold homogeneous database. Our protocol offers enhanced privacy with respect to the current leading K and C protocol simplicity, more efficient in terms of communication rounds, communication cost and computational cost. In our problem, the inputs are the partial databases and the required output is the list of association rules that hold in the unified database with support and confidence no smaller than the given thresholds s and c , respectively.

EXISTING SYSTEM:

Kantarcioglu and Clifton studied that problems and devised a protocol for its solution. The main part of the protocol is a sub-protocol for the secure computation of the union of private subsets that are held by the different players. (The private subset of a given player, as we explain below, includes the item sets that are s -frequent in his partial database. That is the most costly part of the protocol and its implementation relies upon cryptographic primitives such as commutative encryption, oblivious transfer, and hash functions. This is also the only part in the protocol in which the players may extract from their view of the protocol information on other databases, beyond what is implied by the final output and their own input.

While such leakage of information renders the protocol not perfectly secure, the perimeter of the excess information is explicitly bounded and it is argued there that such information leakage is innocuous, whence acceptable from a practical point of view.

DISADVANTAGES OF EXISTING SYSTEM:

- Insufficient security, simplicity and efficiency are not well in the databases, not sure in privacy in an existing system.
- While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players.
- Our protocol may leak is less sensitive than the excess information leaked by the protocol.

PROPOSED SYSTEM:

In Proposed System, propose an alternative protocol for the secure computation of the union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer (what simplifies it significantly and contributes towards much reduced communication and computational costs). While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol.

ADVANTAGES OF PROPOSED SYSTEM:

- We proposed a protocol for secure mining of association rules in horizontally distributed databases that improves significantly upon the current leading protocol in terms of privacy and efficiency.
- The main ingredient in our proposed protocol is a novel secure multi-party protocol for computing the union (or intersection) of private subsets that each of the interacting players holds.

IMPLEMENTATION:

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Modules:

1. Privacy Preserving Data Mining:

Previous work in privacy preserving data mining has considered two related settings. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold. In the first setting, the goal is to protect the data records from the data miner. Hence, the data owner aims at anonymizing the data prior to its release. The main approach in this context is to apply data perturbation. The idea is that. Computation and communication costs versus the number of transactions N the perturbed data can be used to infer general trends in the data, without revealing original record information.

In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multiparty computation. The usual approach here is cryptographic rather than probabilistic. Lindell and Pinkas showed how to securely build an ID3 decision tree when the training set is distributed horizontally. Lin et al. discussed secure clustering using the EM algorithm over horizontally distributed data.

The problem of distributed association rule mining was studied in the vertical setting, where each party holds a different set of attributes, and in [18] in the horizontal setting. Also the work of [26] considered this problem in the horizontal setting, but they considered large-scale systems in which, on top of the

parties that hold the data records (resources) there are also managers which

2. Distributed Computation:

We compared the performance of two secure implementations of the FDM algorithm. In the first implementation (denoted FDM-KC), we executed the unification step using Protocol UNIFI-KC, where the commutative cipher was 1024-bit RSA in the second implementation (denoted FDM) we used our Protocol UNIFI, where the keyed-hash function was HMAC [4]. In both implementations, we implemented Step 5 of the FDM algorithm in the secure manner that was described in later. We tested the two implementations with respect to three measures:

- 1) Total computation time of the complete protocols (FDMKC and FDM) over all players. That measure includes the Apriori computation time, and the time to identify the globally s -frequent item sets, as described in later.
- 2) Total computation time of the unification protocols only (UNIFI-KC and UNIFI) over all players.
- 3) Total message size. We ran three experiment sets, where each set tested the dependence of the above measures on a different parameter:
 - N — the number of transactions in the unified database,

3. Frequent Itemsets:

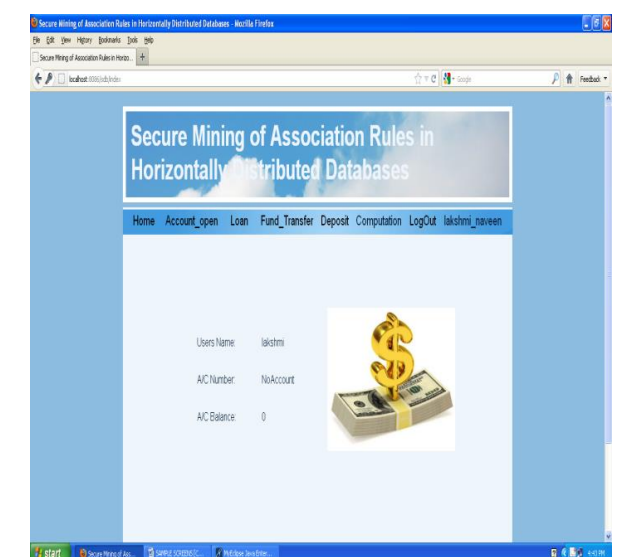
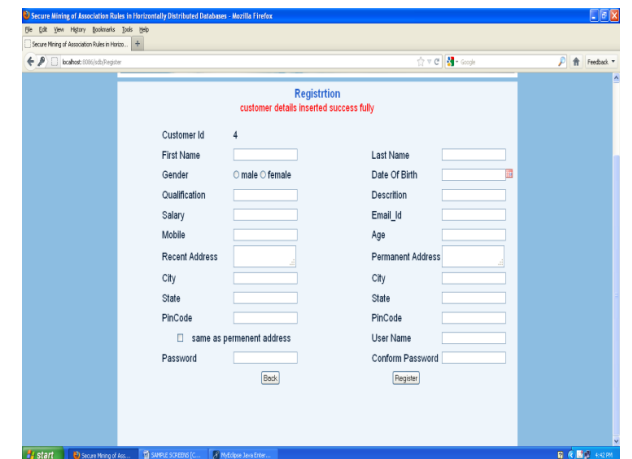
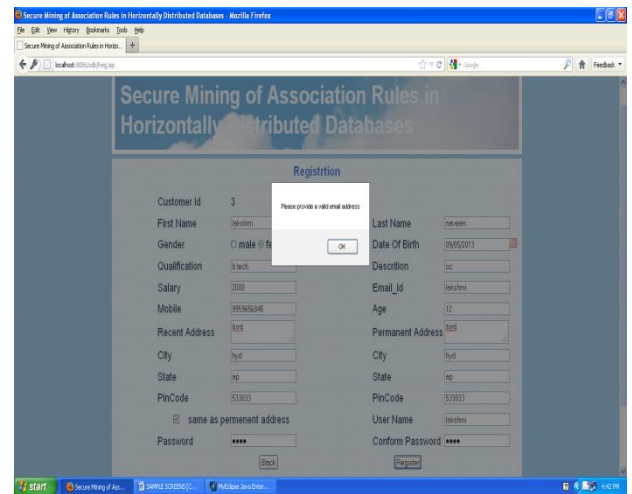
We describe here the solution that was proposed by Kantarcioglu and Clifton. They considered two possible settings. If the required output includes all globally s -frequent item sets, as well as the sizes of their supports, then the values of $\Delta(x)$ can be revealed for all $x \in C_k$. In such a case, those values may be computed using a secure summation protocol (e.g. [6]), where the private addend of P_m is $\text{supp}_m(x) - sN_m$. The more interesting setting, however, is the one where the support sizes are not part of the required output. We proceed to discuss it.

4. Association Rules:

Once the set F_s of all s -frequent itemsets is found, we may proceed to look for all (s, c) -association rules (rules with support at least sN and confidence at least

c), as described in [18]. For $X, Y \in F_s$, where $X \cap Y = \emptyset$, the corresponding association rule $X \Rightarrow Y$ has confidence at least c if and only if $\text{supp}(X \cup Y) / \text{supp}(X) \geq c$, or, equivalently, $CX, Y := M \sum_{m=1}^M (\text{suppm}(X \cup Y) - c \cdot \text{suppm}(X)) \geq 0$. (10) Since $|CX, Y| \leq N$, then by taking $q = 2N+1$, the players can verify inequality (10), in parallel, for all candidate association rules, as described in Section 3. In order to derive from F_s all (s, c) -association rules in an efficient manner we rely upon the following straightforward lemma.

Screen shots



Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

New Account

Account Id 102
 Account Number 42521
 User Name lakshmi
 Account Type
 Branch
 Amount
 Date 04-10-2013

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

Cash Deposit

Deposit NO 13
 Account No 42521
 User Name lakshmi
 Date 04-10-2013
 Branch K.P.H.B
 Deposit Rs
 Balance 20000

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

New Account

Account Id 102
 Account Number 42521
 User Name lakshmi
 Account Type
 Branch K.P.H.B
 Amount 20000
 Date 04-10-2013

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

Cash Deposit

Deposit NO 13
 Account No 42521
 User Name lakshmi
 Date 04-10-2013
 Branch K.P.H.B
 Deposit Rs 1000
 Balance 20000

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

customer account creat success fully

New Account

Account Id 103
 Account Number 42542
 User Name lakshmi
 Account Type
 Branch
 Amount
 Date 04-10-2013

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

customer Deposit success full

Cash Deposit

Deposit NO 14
 Account No 42521
 User Name lakshmi
 Date 04-10-2013
 Branch K.P.H.B
 Deposit Rs
 Balance 21000.0


user details

13	42521	lakshmi	2013-10-04 16:45:40.0	K.P.H.B	1000	20000.0	21000
----	-------	---------	-----------------------	---------	------	---------	-------

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

Fund Transfer



Transfer No: 15

Account No: 42521

User Name: lakshmi

Receiver A/c No: 123456

Receiver Name: naveen

Transfer Rs: 1000

Balance Rs: 21000.0

Branch: K.P.H.B


Date: 04-10-2013

[View](#) [New](#) [Transfer](#)

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

Loan Request



Loan NO: 12

Account No: 42521

User Name: lakshmi

Gender: ☒ male ☐ female

Age: <25

Are You Student: ☒ Yes ☐ No

Income: Medium

Type of Loan: Education

Loan Amount: 100000

Date: 04-10-2013

Branch: K.P.H.B

amount: 20000.0


[Apply](#)

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

customer Deposit success full

Fund Transfer



Transfer No: 16

Account No: 42521

User Name: lakshmi

Receiver A/c No:

Receiver Name:

Transfer Rs:

Balance Rs: 20000.0

Branch: K.P.H.B

Date: 04-10-2013

[View](#) [New](#) [Transfer](#)

user details


Transfer No	A/c No	User Name	Receiver A/c No	Receiver Name	Transfer Rs	Balance	Branch	Date
15	42521	lakshmi	123456	naveen	1000	21000.0	K.P.H.B	2013-10-04 16:46:56.0

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

customer Deposit success full

Loan Request



Loan NO: 13

Account No: 42521

User Name: lakshmi

Gender: ☒ male ☐ female

Age: -select Age-

Are You Student: ☒ Yes ☐ No

Income: -select Income-

Type of Loan: -select Loan-

Loan Amount:

Date: 04-10-2013

Branch: K.P.H.B

amount: 20000.0

[Apply](#)

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

Loan Request



Loan NO: 12

Account No: 42521

User Name: lakshmi

Gender: ☒ male ☐ female

Age: -select Age-

Are You Student: ☒ Yes ☐ No

Income: -select Income-

Type of Loan: -select Loan-

Loan Amount:

Date: 04-10-2013

Branch: K.P.H.B

amount: 20000.0

[Apply](#)

Secure Mining of Association Rules in Horizontally Distributed Databases Mozilla Firefox

Home Account_open Loan Fund_Transfer Deposit Computation LogOut lakshmi_naveen

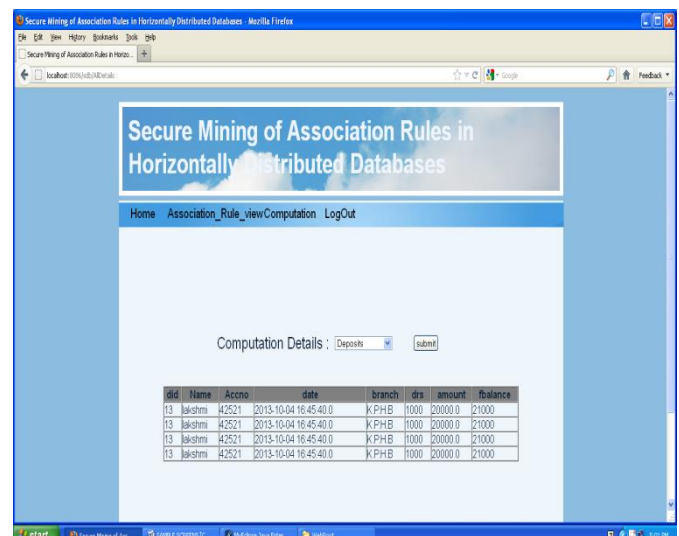
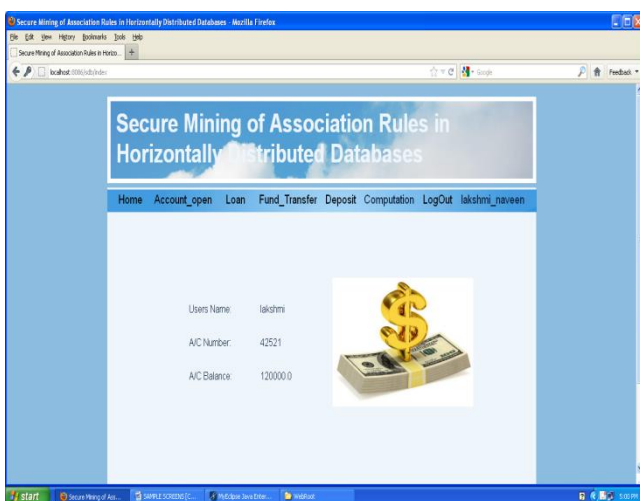
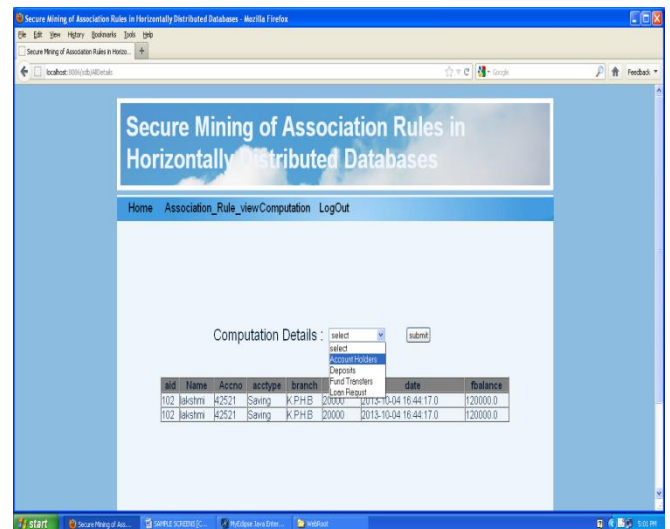
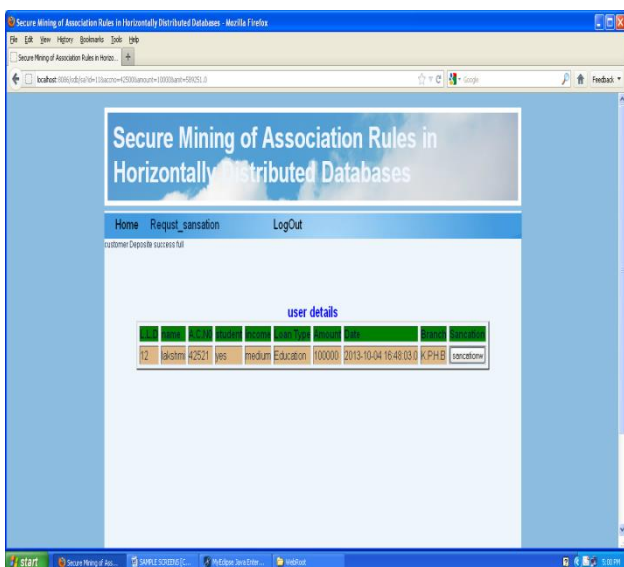
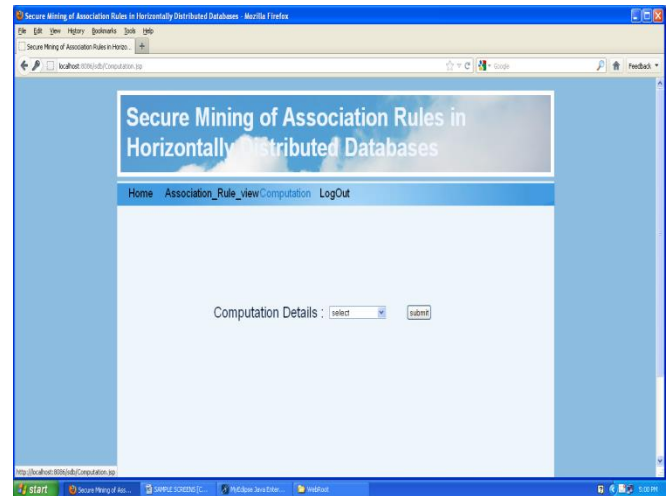
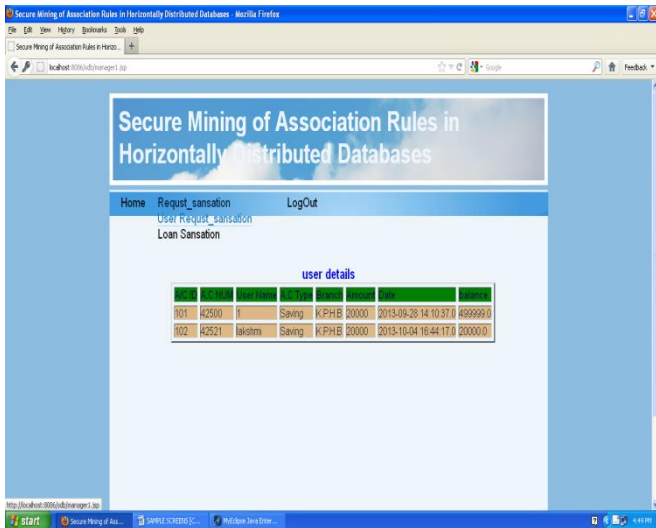
Login!



User Name:

Password:

[Login](#) [register](#)



Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Computation Details : Fund Transfers

Sd	Name	Accno	receiver	name	transfer	balance	branch	/balance	Date
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0
15	lakshmi	42521	123456	haveen	1000	21000	K.PHB	20000	2013-10-04 16:46:56.0

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Itemset Examples

```

graph TD
    Start([Start]) --> GetFrequent[Get Frequent Items]
    GetFrequent --> GenCandidate[Generate Candidate Itemsets]
    GenCandidate --> GetFrequent
    GetFrequent --> GenStrong[Generate Strong Rules]
    GenStrong --> GenStrong
  
```

useritem Examples

A
B
C
D
E
F
G
H
I
J

View Item Set

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Computation Details : Loan Request

Ld	Name	Accno	Gender	Age	Student	Income	LoanType	Amount	Date	Branch
10	T	42500	male	>25	yes	medium	Education	10000	2013-09-28 16:39:27.0	K.PHB
11	T	42500	male	<25	yes	low	Education	10000	2013-09-30 14:34:56.0	K.PHB
12	lakshmi	42521	male	<25	yes	medium	Education	100000	2013-10-04 16:46:03.0	K.PHB

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Generated Set = Null

Generate Strong Rules

ITEM LIST	LOAN ITEM LIST	FIRST ITEMSET	APRIORI ALGORITHM SET	ASSOCIATION RULE	MIN ASSOCIATE RULE
Min support: 2 Min confidence: 1 <input type="button" value="Apply"/>	A, C, F, B B, C, G, 2 A, B, G, 2 B, C, G, 1 A, C, G, 1 B, C, F, B A, B, G, 2 A, B, G, 1 B, C, G, 1				

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Itemset Examples

```

graph TD
    Start([Start]) --> GetFrequent[Get Frequent Items]
    GetFrequent --> GenCandidate[Generate Candidate Itemsets]
    GenCandidate --> GetFrequent
    GetFrequent --> GenStrong[Generate Strong Rules]
    GenStrong --> GenStrong
  
```

useritem Examples

View Item Set

Secure Mining of Association Rules in Horizontally Distributed Databases - Mozilla Firefox

Secure Mining of Association Rules in Horiz...

Secure Mining of Association Rules in Horizontally Distributed Databases

Home Association_Rule_viewComputation Logout

Generated Set = Null

Generate Strong Rules

ITEM LIST	LOAN ITEM LIST	FIRST ITEMSET	APRIORI ALGORITHM SET	ASSOCIATION RULE	MIN ASSOCIATE RULE
Min support: <input type="text"/> Min confidence: <input type="text"/> <input type="button" value="Apply"/>	A, C, F, B B, C, G, 2 A, B, G, 2 B, C, G, 1 A, C, G, 1 B, C, F, B A, B, G, 2 A, B, G, 1 B, C, G, 1	[A]=5 [B]=6 [C]=8 [D]=3 [E]=9 [F]=2 [G]=4 [H]=2 [I]=4 [J]=3	[A, B]=1 [A, C]=2 [A, D]=3 [A, E]=4 [A, F]=1 [A, G]=2 [B, C]=1 [B, D]=2 [B, E]=1 [B, F]=2 [B, G]=2 [C, D]=1 [C, E]=2 [C, F]=1 [C, G]=2 [D, E]=1 [D, F]=2 [D, G]=2 [E, F]=1 [E, G]=2 [F, G]=2	[A, B]--> [C] [A, C]--> [B] [A, D]--> [E] [A, E]--> [D] [A, F]--> [G] [A, G]--> [F] [B, C]--> [D] [B, D]--> [C] [B, E]--> [F] [B, F]--> [E] [B, G]--> [C] [C, D]--> [E] [C, E]--> [D] [C, F]--> [G] [C, G]--> [F] [D, E]--> [F] [D, F]--> [E] [D, G]--> [C] [E, F]--> [D] [E, G]--> [C] [F, G]--> [C]	[B, G, 2]--> [C] [A, 2]--> [B] [A, 2]--> [C] [A, 2]--> [D] [A, 2]--> [E] [A, 2]--> [F] [A, 2]--> [G] [A, 2]--> [H] [A, 2]--> [I] [A, 2]--> [J] [A, 2]--> [K] [A, 2]--> [L] [A, 2]--> [M] [A, 2]--> [N] [A, 2]--> [O] [A, 2]--> [P] [A, 2]--> [Q] [A, 2]--> [R] [A, 2]--> [S] [A, 2]--> [T] [A, 2]--> [U] [A, 2]--> [V] [A, 2]--> [W] [A, 2]--> [X] [A, 2]--> [Y] [A, 2]--> [Z]

CONCLUSION:

In this project we devise a protocol for secure mining of association rules in horizontally partitioned distributed databases. The protocol is more efficient than current leading K and C protocol. The main ingredients of this protocol are two novel secure multiparty algorithms in which these two main operations are union and intersection. The protocol exploits the fact that the underlying problem is of interest only if the number of player is more than two. The direction to future work is to devise an efficient protocol for inequality verifications that uses the existence of semi-honest third party and another in Implementation of the techniques to the problem of distributed association rule mining in vertical setting.

REFERENCES:

- [1] Tamirtassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE transactions on knowledge and data engineering, 2013.
- [2] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data

Mining, Edmonton, Alberta, Canada, July 23-26 2002, pp. 639–644.

- [3] M.Kantarcioglu and C. Clifton., "Privacy-preserving distributed mining of association rules on horizontally partitioned data", IEEE Transactions on Knowledge and Data Engineering, 16:1026–1037, 2004.
- [4] R.Agrawal and R. Srikant., "Privacy-preserving data mining", SIGMOD Conference, pages 439–450, 2000.
- [5] A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules", In KDD, pages 217–228, 2002.
- [6] M. Kantarcioglu, R. Nix, and J. Vaidya, "An efficient approximate protocol for privacy-preserving association rule mining", In PAKDD, pages 515– 524, 2009.
- [7] M. Freedman, Y. Ishai, B. Pinkas, and O. Reingold., "Keyword search and oblivious pseudorandom functions", In TCC, pages 303–324, 2005